大和大学　研究紀要　第６巻　政治経済学部編　2020年3月　pp.85～94

ベイジアン・アプローチによるNash均衡への収束と均衡選択

Bayesian Approach to Convergence to Nash Equilibrium and Equilibrium Selection

吉　川　　満
KIKKAWA Mitsuru

要　　旨
本稿では情報不完備なゲームにおいて，各プレイヤーがベイズ学習を行い，自身の戦略を更新していく場合における Nash 均衡への収束について考察した。特にここではマルチンゲールを用いて，各プレイヤーの戦略が Nash 均衡へ収束することを示した。また，当初複数の純粋戦略の Nash 均衡があるコーディネーションゲームにおいて均衡選択を行った。その結果，純粋戦略の数が2つの場合，既存の確率的進化ゲーム理論と同様の結論，均衡が一意になることが見出され，戦略の数が十分大きい場合には，Random 行列の議論を適用し，均衡選択について考察した。

Abstract
In this paper, we examine the game in which each player chooses one's own strategy by Bayesian learning. We then prove that the sequence of each player's expected utility leads to Nash equilibrium with a martingale. This paper examines equilibrium selection under the likelihood in two distributions: binomial and normal distributions. In the binomial distribution, this equilibrium selection is similar to a stochastic evolutionary game. In the normal distribution, this equilibrium selection is used with random matrix theory to derive some conditions.

キーワード：情報不完備ゲーム，ベイズ学習，進化ゲーム理論，コーディネーションゲーム
keywords：Game with Incomplete Information, Bayesian Learning, Evolutionary Game Theory, Coordination Game

1.　はじめに

本稿ではプレイヤーが対戦相手の過去の行動やゲームのルールを含む外部の情報をベイズ学習により，戦略を変更していくという情報不完備ゲームにおいてNash均衡への収束性とコーディネーションゲームにおける均衡選択について考察した。

今まで各プレイヤーがベイズ学習して戦略を変更していくという研究として Jordan [4, 5], Kalai and Lehrer [6], Nachbar [13], Nyarko [14]等がある。これらの研究では対戦相手の戦略に対して，過去の行動から生成される今期の予想の度合いを示す，信念を持ち，この信念がベイズの公式で学習することにより，無限回の繰り返しゲームにおいて，Nash均衡への収束することを示している。

Kalai and Lehrer [6]，Nyarko [14]では信念の絶対連続性の下で，プレイヤーの戦略の分布関数がNash均衡へ収束することを証明している。そこではNash均衡が混合戦略のみのようなゲームにおいては，信念の絶対連続性の条件を満たさないため，このベイズの公式を使った手法では一般的には成り立たない。そこでNachbar [13]はこの問題を異なる条件の下で解決した[1]。一方 Jordan [4]では各プレイヤーがどのようなゲームを行っているのか分からない，つまり共有知識がない場合のゲームにおいても，均衡状態において，プレイヤーの戦略の分布関数がNash均衡へ収束することを示している。

本稿では，既存の研究にあるようなプレイヤーの戦略の分布関数の収束性を考察するのではなく，期待利得の列からNash 均衡への収束することを示した。ここでは各プレイヤーが毎期，相手の行動やゲームのルールなどの外部からの情報を得て，ベイズ学習を行い，各自の戦略を決定する。例えばある寡占市場において，外部からの情報を参考にして，

*大和大学政治経済学部　令和元年12月11日受理

[1] 一連の研究に関しては，Fudenberg and Levine [1]を参照されたい。

ライバル企業の費用関数を予測し，ライバル企業の利得を計算し，供給量・戦略を予測し，自らの供給量・戦略を決定する。長期のゲームにおいては途中で情報を参考に戦略を変更していくということは自然な意思決定方法である[2]。

本稿は次のように構成されている。第 2 節では，ゲームを定式化し，Nash 均衡への収束性を示す。また具体例として，不完備情報がある Cournot 複占市場を取り上げる。第 3 節では，尤度がある分布に従っている場合における均衡選択を考察する。第 4 節では，まとめる。

2. 定式化と主定理

ここでは無限回繰り返しゲームにおいて，毎期に各プレイヤーは相手の行動や外部情報等を参考にし，相手の戦略を推論しながら，自らの戦略を変更，決定していくという情報不完備ゲームを想定している。ここでの情報不完備ゲームは，岡田 [15]を参考にして，定式化した。

まず各プレイヤーは対戦相手の戦略に対して，過去の行動等から生成される今期の予想の度合いを示す，信念を持つ。各プレイヤーは，この信念をもとにし，受け取った外部情報から自らの利得が最大になるように戦略を決定する。プレイヤーがこの信念を持つとすると，ベイズ学習で戦略を採用する確率が定まると考える。このようなベイジアンモデルは，特定の事象に関して，条件付き確率を用いて，対戦する相手に対する条件付き確率を推定することに相当する[3]。このようなことを無限回繰り返し，均衡状態が存在すれば，そこにおいてはこのような推論の連鎖が停止する。この推論が停止する均衡状態を調べた。

定義 1 情報不完備ゲームは，組$G=(N,\{S_i,C_i,u_i,p_i\}_{i\in N})$ で定義される。
ただし，$N=\{1,\cdots,n\}$はプレイヤーの集合，S_iはプレイヤーiの行動の集合，C_iはプレイヤーiのタイプの集合，u_iは直積集合$S_1\times\cdots\times S_n\times C_1\times\cdots\times C_n$上の実数値関数で，プレイヤー$i$の利得関数を表す。$p_i$は各$c_i\in C_i$ に対して直積集合$C_1\times\cdots\times C_{i-1}\times C_{i+1}\times\cdots\times C_n$ 上の同時確率分布$p_i(c_1,\cdots,c_{i-1},c_{i+1},\cdots,c_n|c_i)$を対応させる。

仮定　プレイヤー$i(=1,\cdots,n)$のタイプ集合C_iは有限集合とする。

情報不完備ゲームにおけるプレイヤーiの戦略π_iとは，タイプ集合C_iから行動集合S_iへの写像である。すべてのプレイヤーの戦略の組$\pi=(\pi_1,\cdots,\pi_n)$に対して，プレイヤーiの条件付き期待利得を$Eu_i(\pi|c_i)=\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{c_j\in C_j}u_i(\pi_1(c_1),\cdots,\pi_n(c_n);c)p_i(c_{-i}|c_i)$と定義することができる。

---

2 繰り返しコーディネーションゲームを行う実験（Huyck, et al. [3]）では，毎期どのくらいの人が Pareto 最適な行動を採用していないのかという割合を公表し，次のゲームを行うと Pareto 劣位な均衡に達成してしまうという実験結果を見出している。このことからも自らの戦略・行動を決定する際に，他者の行動を利用していることが分かる。

3 ベイズの定理とは，$A,B\in\Omega$を事象とし，事象 Bが起きた後，事象 Aの条件付き確率は$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(B|A)P(A)}{P(B)}$ が成り立つことを言う，ただし$P(A)$は Aの事前確率とする。$P(A|B)$は事象 Bが起きた後，事象 Aの条件付き確率であり，事後確率と呼ばれる。$P(B|A)$は事象 Aが起きた後，事象 Bの条件付き確率であり，尤度と呼ばれる。$P(B)$は Bの事前確率であり，規格化定数である。

定義 2　情報不完備ゲームの戦略の組$\pi^* = (\pi_1^*, \cdots, \pi_n^*)$が G のベイジアン均衡点であるとは，すべてのプレイヤー$i(=1,\cdots,n)$のすべてのタイプ$c_i \in C_i$とすべての戦略π_iに対して，次が成り立つことである。

$$Eu_i(\pi^*|c_i) \geq Eu_i(\pi_i, \pi_{-i}^*|c_i)$$

ただし，右辺はプレイヤーiだけが戦略をπ_i^*からπ_iに変更する場合のプレイヤーiの条件付き期待利得を表す。

定理　各プレイヤーが相手の行動に関する信念についての最適な戦略は Nash 均衡へ収束する。

証明　付録 A

この定理から今までの結果と同様に，各プレイヤーがベイズ学習を用いて，戦略を採用するようなゲームにおいて，均衡状態において Nash 均衡へ収束することが分かった。ただし既存の研究とは異なり，ここでは期待利得の列の収束性を考えているために，今までの研究からでは条件を満たさなかった Nash 均衡が混合戦略のみの場合であっても，この結果が成り立つ。例えば Replicator 方程式を用いた進化ゲーム理論[4]においては，マッチングペニーゲームのような混合戦略のみが Nash 均衡となるゲームにおける戦略を採用する確率の変動はリミットサイクルとなり，期待利得は一定であるということが知られている。これからも分かるように，プレイヤーの戦略の分布関数の収束性ではなく，期待利得の列の収束を考えることにより，Kalai and Lehrer [6]の研究における困難を別の視点で考察した。

例　情報不完備なクルノー複占市場（岡田 [15][5]）

ライバル企業の費用関数に関して，各企業はベイズ学習を行いながら，自企業の戦略を決定していくとする。このとき，ゲームの均衡は Cournot-Nash 均衡へ収束する。詳細については付録 B を参照されたい。

3. 均衡選択

前節では尤度が一般的な場合における基本的な性質を調べ，均衡状態においては Nash 均衡へ収束することが分かった。本節では尤度が特定の分布に従っている場合における均衡選択を行う。

3.1 2 戦略：2 項分布

ここでは様々な情報を受け取ったとしても，それに従い戦略を変更するのか，しないのかという 2 通りの選択肢がある場合を考える。そのため仮に戦略 A を採用したプレイヤーがいたとしても，その情報を利用していないプレイヤーが存在することがある。よって戦略 A を採用する確率を p とすると，ここでの尤度，$P(s_2|A)$は 2 項分布 ${}_n\mathrm{C}_y p^y (1-p)^{n-y}$に従うということが分かる。通常プレイヤーの利得が低くなるような戦略を採用しないが，ここでは突然変異戦略[6]としてこのような戦略を採用するとする。またこの突然変異戦略を採用する確率を ε で表す。

---

[4] Replicator 方程式はプレイヤーの学習プロセスを説明するもので，Nash 均衡まで収束性を考察することができる。

[5] 自社のみの費用関数を知っている場合とライバル企業の費用関数をも知っている場合を比較し，情報構造が企業行動に与える影響を調べ，不確実性がある場合には企業にとって不利であるということを導いている。

[6] 各時点においてある確率で本来選ぶべき戦略とは異なる戦略を，何かの間違いなどで選ぶ場合がある。このような戦略を突然変異戦略と呼ぶ。

今までこのような状況を定式化した研究として，Kandori, et al. [7]等に代表される確率的進化ゲーム理論がある。ただしこれらはある単調な決定論的な関数，例えば $z_{t+1} = b(z_t)$に従い，戦略を採用する確率・プレイヤーが増加するような理論を考察していた。本稿では各プレイヤーはベイズ学習によって戦略を選択するとする。

よってここで各プレイヤーは様々な情報を受けた後, 戦略 A を採用する確率はベイズの公式を次のように表すことができる。

$$\mathrm{P}(A|s_2) = Z^{-1}\mathrm{P}(A)\ {}_n\mathrm{C}_y p^y (1-p)^{n-y} \tag{1}$$

Z^{-1}は規格化定数，n はプレイヤーの数，y は戦略を変更したプレイヤーの数を表している。

	A	B
A	a,a	0,0
B	0,0	b,b

利得表 1　(a > b > 0)

ここで利得表 1 のようなコーディネーションゲームにおいて，このゲームのベイジアン均衡点を考える。

命題 1　プレイヤーの戦略に突然変異がある場合の対称 2 人コーディネーションゲームにおいて，長期均衡においては突然変異戦略の大きさにより一意の均衡点，戦略の組 (戦略 A，戦略 A) のみがベイジアン均衡点となる。

証明　付録 A

この命題 1 から戦略に突然変異がある場合, 複数均衡あったゲームにおいて, 均衡が一意になるということが分かった。またこの結論は Kandori, et al. [7]等における確率的進化ゲーム理論においても，利得に関して単調性があるため，同様の結論が得られる。

3.2 r 戦略：正規分布

次に戦略の数 r が十分大きい一般的な場合を考察する。先ほどの特に 2 項分布は試行回数 (ここでは戦略の数) r が十分大きい場合には，正規分布に従うことが知られている。よって戦略の数 r が十分大きい場合には(1)式は次のように式変形することができる。

$$P(A|s_j) = Z'^{-1}P(A)exp\left(\frac{-(y-m)^2}{2\rho^2}\right), \quad j = 1, \cdots, r$$

Z'^{-1}は規格化定数，$m > 0$ は y の平均，$\sigma^2 < \infty$は y の分散を表している。よってこの式は情報を受け取った場合の戦略を採用する確率は情報を受け取る前の確率にホワイトノイズがある場合と解釈することができる。これを中心極限定理を用いて，整理すると，次を得る。

$$Z'^{-1}\frac{P(A|s_j)}{P(A)} = \sum_{j=1}^{r} \xi_j \coloneqq \xi$$

ただし ξは上記の条件を満たすようなホワイトノイズとする。よって情報を受け取った後，戦略を採用する確率 (事後分布) は，事前分布がノイズによって変化・ゆらぐということを表している。またこのような状況の時，戦略の分布に関

してまとめると次の補題を得る。

補題1 戦略jを採用する確率 p_jは正規分布に従う。

この補題1から分散の大きさにより，Nash均衡へ到達可能(approchable)であることが分かる[7]。また推移行列の要素が正規分布に従っているr行r列の推移行列の性質を用いると，安定性に関して次のことが分かる[8]。

命題2　この推移行列は実対称の Random 行列[9]となり，この行列において内点が局所安定，つまり混合戦略が進化的に安定な戦略(ESS)となるための十分条件は，最大固有値$2y' < 1$のときである。またある純粋戦略が ESS となるための十分条件は，最大固有値$2y' > 1$のときである。

証明 付録A

この命題2から尤度の分散によって，どの均衡へ収束するのかが分かった。この尤度の分散を次の補題を利用して，特徴づける。

補題2(Losert and Akin [10])　平均利得の上昇率は，各プレイヤーの戦略の分散に等しい。また平均利得は単調増加する。

証明 付録A

この補題は生物学では Fisher の基本定理と呼ばれており，この補題から戦略について戦略の多様性があればあるほど，平均利得は上昇する(Kikkawa [9])。前節までにおいて，ベイズ学習はNash均衡への収束していく動学プロセスを示していた。そのため当初戦略が不均一で，戦略に関して多様性がある場合，時間とともに戦略が徐々にNash均衡へ収束，均一化していくため，当初は平均利得の上昇率が高くなるためである。

以上の2つの結果から情報に関して，その情報の分散の大きさに依存していた。まず分散が小さい場合には，戦略を変更したときの利得が小さいために，戦略分布は平均の近傍にあり，均衡戦略は混合戦略，内点解となりやすい。一方分散が大きい場合には，戦略を変更したときの利得が大きいために，戦略分布は裾野が広いが，時間とともに均衡戦略は純粋戦略となりやすいということが分かる。繰り返しコーディネーションゲームの実験においても見出されているよ

[7] Harsanyi [2]は混合戦略の解釈を行う際に，各プレイヤーの利得がノイズによりランダムに変動し，ノイズが小さくなる際に，Nash均衡へ到達する場合を到達可能(approchable)という概念を提示した。Kikkawa [9]では Replicator 方程式を用いた進化ゲーム理論の枠組みで，到達可能について諸性質を見出している。

[8] この手法の詳細については，付録Cを参照されたい。

[9] Random 行列とは，行列の要素が確率分布に従う確率変数として与えられると仮定する行列モデルのことを言う。この Random 行列理論は，物理学の他，生物学，金融工学等様々な分野に応用されている。例えば，Kikkawa [8]では，各プレイヤーがランダムにマッチしたプレイヤーとゲームをプレイするということをランダム行列理論を適用し，進化ゲーム理論を定式化している。

うに，利得が多く得られる，抜け駆けの誘因が大きいときには，Nash 均衡はどちらか一方の純粋戦略となり，コンフリクトが解消されやすいということが示唆されており，この実験結果と整合的である (Huyck，et al. [3])。

4. まとめ

以上のように一般的なゲームにおいて，各プレイヤーがベイズ学習を行い戦略を採用していくゲームを考察した。特にここではマルチンゲールを用いて各プレイヤーの戦略が Nash 均衡へ収束することを示した。また戦略の数が 2 つの場合，当初複数の純粋戦略の均衡がある対称 2 人コーディネーションゲームにおいて均衡選択を行った。その結果既存の確率的進化ゲーム理論と同様の結論，突然変異戦略の大きさにより，均衡が一意になることが見出された。また戦略の数が十分大きい場合には，Random 行列の議論が適用でき，これから均衡選択をすることができた。その結果情報の分散の大きさに依存しており，大きい場合には利得を多く得る機会があることからどちらか一方の純粋戦略となり，コンフリクトが解消し，小さい場合には混合戦略となり，解消しないことが分かった。

付録 A　証明

定理の証明　ここで t 期の期待利得を$u_i(\pi) \coloneqq Y_t, c_i = F_t$とする。ここでベイズの公式から t 期の期待利得は次のように表すことができる。

$$E(Y_{t+1}|F_t) = E(X_t)E(Y_t|F_t) = E(X_t)Y_t = \mu Y_t$$

ただし Xを尤度とし，μ は尤度の平均とする。これから次を得る。

$$E\left(\frac{Y_{t+1}}{\mu^{t+1}}\middle|F_t\right) = \frac{Y_t}{\mu^t}$$

よって期待利得を尤度の平均で割ったものはマルチンゲールであることが分かる。またマルチンゲールの収束定理[10]が成り立ち，均衡状態は Nash 均衡であるので，この期待値の列は Nash 均衡へ収束する。

命題 1 の証明　s_1, s_2をそれぞれ戦略 A，B を採用している状態とする。また前期戦略 A を採用しており，今期戦略 B を採用する確率を$P(B|s_1)$，逆に前期戦略 B を採用しており，今期戦略 A を採用する確率を$P(A|s_2)$とするとベイズの公式から，次のように定式化することができた。

$$\mathrm{P}(B|s_1) = Z^{-1}\mathrm{P}(B)\ {}_n\mathrm{C}_y p^{n-y}(1-p)^y,$$

$$\mathrm{P}(A|s_2) = Z^{-1}\mathrm{P}(A)\ {}_n\mathrm{C}_y p^{y}(1-p)^{n-y}.$$

均衡状態において，このゲームの推移行列は次を満たしている。

$$\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} 1-P(A|s_2) & P(A|s_2) \\ P(B|s_1) & 1-P(B|s_1) \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

推移行列において状態s_1にいる確率は$1-P(A|s_2)$の確率で戦略を変更しなかった場合と，$P(A|s_2)$の確率で状態s_2から戦略を変更した場合で定式化される。

また状態s_2の場合も同様である。この推移行列から均衡状態においては，

$$\frac{s_2}{s_1} = \frac{P(B|s_1)}{P(A|s_2)}$$

[10] X_nは数列であり，$sup(X_n) < \infty$であるサブマルチンゲールとするとき，$n \to \infty$，X_nは$E(|X|) < \infty$ である，$\lim X$へ概収束(a.s.収束)する。

を満たすことが分かる。よってこれを満たすような関係式は次のようになる。

$$\frac{s_2}{s_1}=\frac{P(B|s_1)}{P(A|s_2)}=\frac{\varepsilon^{N-N^*}(1-\varepsilon)^{N^*}+O\left(\varepsilon^{N-N^*+1}\right)}{\varepsilon^{N^*}(1-\varepsilon)^{N-N^*}+O(\varepsilon^{N^*+1})}$$

これから$\varepsilon \to 0$のとき，$\frac{s_2}{s_1}$は 0 に収束する。よって均衡状態においては，$s_2 \to 0$となる。すなわち情報に従わないプレイヤーが少なければ，両プレイヤーが共に戦略 A を採用する状態がベイジアン均衡点となる。

命題 2 の証明　上述の仮定において推移確率が平均 m，分散 y という正規分布に従っている。またゲームは対称 2 人ゲームであるので，$a_{ij}=a_{ji}$ （ただし a_{ij},a_{ji}は推移行列の要素とする) となる実対称行列であると分かる。よってこの推移行列は実対称の Random 行列を用いて，記述することができる。

ここで推移行列の固有値とベイズの公式のアナロジーを利用する。ここではベイズの公式の尤度は固有値の分布に対応することが分かる。詳しくは付録 C を参照されたい。このときの推移行列の固有値の分布は，Wigner の半円則[11]に従うことが知られている。よってこの半円則を利用すると，最大固有値$2y'$であることが知られている。これから内部均衡点の局所安定性の十分条件を考察すると，$2y'<1$の場合，この内部均衡点，混合戦略は局所安定であり，ESS となることが分かる。またこの裏を取れば，純粋戦略の場合も分かる。

補題 2 の証明　定理からこの期待利得の列はマルチンゲールであり，単調増加列であったので，利得の単調増加性が存在する。

次に利得の単調増加性が存在する場合には，戦略を採用する確率の変動は進化ゲーム理論で使用される，次のような Replicator 方程式が導かれることが知られている。詳しくは Metcalfe [12]等を参照されたい。

$$\dot{x}_j=x_j\left(g_j-\bar{g}\right)$$

ただしg_jを戦略 j を採用した場合の利得，$\bar{g}$を平均利得とする。

この Replicator 方程式からの平均利得$\bar{g}=\sum_{j=1}^{r}x_j g_j$に着目する。これを時間で微分すると，次を得る。

$$\frac{d\bar{g}}{dt}=\frac{d\left(\sum_{j=1}^{r}x_j g_j\right)}{dt}=\sum_{j=1}^{r}\left(\frac{dx_j}{dt}\cdot g_j\right)$$

ここで$\frac{dx_j}{dt}$は Replicator 方程式であるので，次を得る。

$$\frac{d\bar{g}}{dt}=\sum_{j=1}^{r}\left(x_j\left[g_j-\bar{g}\right]\cdot g_j\right)=\sum_{j=1}^{r}x_j\left[g_j-\bar{g}\right]^2 \qquad (\geq 0)$$

を得る。　なぜなら，

$$\sum_{j=1}^{r}x_j\left[g_j-\bar{g}\right]^2=\sum_{j=1}^{r}x_j\left(g_j^2-2g_j\bar{g}+\bar{g}^2\right)=\sum_{j=1}^{r}x_j g_j\left(g_j-\bar{g}^2\right)-\bar{g}\sum_{j=1}^{r}x_j g_j+\bar{g}^2\sum_{j=1}^{r}x_j=\sum_{j=1}^{r}x_j\left(g_j-\bar{g}\right)g_j-\bar{g}^2+\bar{g}^2$$
$$=\sum_{j=1}^{r}x_j\left(g_j-\bar{g}\right)$$

を得る。もちろん最終項は各プレイヤーの戦略の分散である。

[11] Wigner の半円の法則とは，ある一般的な特性を満たす分布した非常に大きな対称実行列の場合，固有値の分布は半円関数となることをいう。詳細については，Mehta [11]等を参照されたい。

付録 B　情報不完備なクルノー複占市場（岡田 [15]）

n 企業が同質な財を生産し，市場に供給している。企業$i(=1,\cdots,n)$の供給量をq_iとすると，財の価格 p は逆需要関数 $p = max\{a - b(\sum_{i=1}^{n} q_i), 0\}$，$a, b > 0$ によって決まるとする。特にここでは企業は生産費用に関して様々なタイプがあり，各企業は自分のタイプは知っているが，ライバル企業のタイプは知らないような状況を考える。各企業$i(=1,\cdots,n)$は費用関数$C_i(q_i) = c_i q_i, 0 \le c_i \le K_i$を持つ。ただし限界費用$c_i$は区間$[0, K_i]$に値をとる確率変数である。企業の限界費用$c_i$の同時確率分布関数を$F(c_1,\cdots,c_n)$とする。

また企業 i の利潤関数は$u_i(q_i, q_{-i}; c_i) = pq_i - c_i q_i$である。企業 i の戦略π_iは，限界費用のあらゆる値$c_i \in [0, K_i]$に対して財の供給量$\pi_i(c_i)$を対応させる関数である。戦略の組$\pi = (\pi_1,\cdots,\pi_n)$に対して，企業 i の限界費用が c_iであるときの企業 i の条件付き期待利潤は次のようになる。

$$Eu_i(\pi_i, \pi_{-i}|c_i) = \int_0^{K_j} u_i(\pi_i(c_i), \pi_{-i}; c_i) dF(c_{-i}|c_i)$$

ただし$F(c_{-i}|c_i)$はc_iが与えられた場合の確率変数c_{-i}の条件付き確率分布関数である。以下では，$E(\cdot), E(\cdot|c_i)$はそれぞれ期待値と条件付き期待値を表す。

命題 B　上記のような Cournot 複占市場$(i = 1,2)$において，均衡状態における企業の戦略の組 (π_1^*, π_2^*)は次の情報完備な Cournot-Nash 均衡へ収束する。

$$(\pi_1^*, \pi_2^*) = \left(\frac{a - 2c_1 + c_2}{3b}, \frac{a - 2c_2 + c_1}{3b}\right)$$

証明　企業$(i = 1,2)$の期待利潤の計算式

$$\mathrm{E}u_i(\pi_i, \pi_j|c_i) = \int_0^{K_j} \left[\left\{a - b\left(q_i + \pi_j(c_j)\right)\right\} q_i - c_i q_i\right] dF(c_j|c_i)$$
$$= \left\{a - c_i - b\left(q_i + E(\pi_j|c_i)\right)\right\} q_i, q_i = \pi_i(c_i)$$

これをq_iに関しての 1 階条件から，$q_i = \frac{a - c_i - bE(\pi_j|c_i)}{2b}$ を得る。よって$i, j = 1,2 (i \ne j)$に対して，均衡状態における企業 i の最適な供給量は次のようになる。

$$\pi_i^*(c_i) = max\left[\frac{a - c_i - bE(\pi_j|c_i)}{2b}, 0\right] \quad (a.e.) \qquad \text{(B.1)}$$

定理からこの条件付き期待利得の列はマルチンゲールであり，マルチンゲールの収束定理から$E(\pi_j|c_i) = \pi_j$が成り立つ。これと(B.1)式から，$q_i = \frac{a - c_i - bq_j}{2b}$，$q_j = \frac{a - c_j - bq_i}{2b}$ という反応関数を得る。この反応関数から各企業の最適な供給量$(\pi_1^*, \pi_2^*) = \left(\frac{a - 2c_1 + c_2}{3b}, \frac{a - 2c_2 + c_1}{3b}\right)$が得られる。これは情報完備な Cournot 複占市場における Cournot-Nash 均衡であった[12]。

[12] 動学を用いて寡占市場を研究した研究に Vega-Redondo [16]がある。しかし Vega-Redondo [16]では戦略の調整プロセスにおいては，ある期に最も利得を得ている戦略を模倣していくという，模倣過程で各企業の戦略を決定するとしていた。そのため均衡状態においては，競争均衡（ワルラス均衡）へ収束することを示した。しかし本稿ではライバル企業の行動を模倣していくのではなく，自企業の利益が最大となるように，戦略を決定している。そのためは Vega-Redondo [16]の結果とは異なるものの，Nash 均衡へ収束する。

付録 C　推移行列のベイズ学習的アプローチ

ここでは推移行列とベイズの公式$P(E|C) = Z^{-1}P(C|E)P(E)$とのアナロジーにより考察する。$B$を推移行列とし，$t$期から $t+1$ 期への推移を次のように表すことができた。

$$B_{t+1} = AB_t$$

ここでこの行列 A の固有値を考察する。固有値とは次を満たすものであった。

$$Au = \lambda u$$

ただし u は右固有ベクトルである。行列 A は n 行 n 列であるので，固有値は n 個あるので，右固有値ベクトルもそれぞれの固有値に応じて，n 個存在する。そのそれぞれを添字を付した記号を用いて，$\lambda_i, u_i \ \ i(= 1, \cdots, n)$と表しておく。

また n 個の右固有ベクトルの一次結合を満足するような n 個のc_iは一意的に存在することは線形代数で知られている。

$$B_t = c_1u_1 + c_2u_2 + \cdots + c_nu_n$$

これらを用いると，次のように変形することができる。

$$\begin{aligned} B_{t+1} = AB_t &= A(c_1u_1 + c_2u_2 + \cdots + c_nu_n) \\ &= c_1Au_1 + c_2Au_2 + \cdots + c_nAu_n \\ &= c_1\lambda u_1 + c_2\lambda u_2 + \cdots + c_n\lambda u_n \end{aligned}$$

つまり固有値の分布$(\lambda_1, \cdots, \lambda_n)$により次期の状態が定まると分かる。そこで推移行列の場合固有値をベイズ統計における尤度として捉えると，ベイズ学習においてもこのような推移行列における固有値の議論ができる。

参考文献

[1] Fudenberg, Drew and David K. Levine, *The Theory of Learning in Games*, MIT Press, 1998

[2] Harsanyi, John C. (1973) "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points," International Journal of Game Theory, Vol. 2, pp. 1-23

[3] Van Huyck,　John B., Raymond C. Battalio and Richard O. Beil (1990) "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure," The American Economic Review, Vol. 80, pp. 234-248

[4] Jordan, J. S. (1991) "Bayesian Learning in Normal Form Games," Games and Economic Behavior, Vol. 3, pp. 60-81

[5] Jordan, J. S. (1995) "Bayesian Learning in Repeated Games," Games and Economic Behavior, Vol. 9, pp. 8-20

[6] Kalai, Ehud and Ehud Lehrer (1993) "Rational Learning Leads to Nash Equilibrium," Econometrica, Vol. 61, pp. 1019-1045

[7] Kandori, Michihiro, George J. Mailath and Rafael Rob (1993) "Learning, Mutation, and Long Run Equilibria in Games," Econometrica, Vol. 61, pp. 29-56

[8] Kikkawa, Mitsuru (2009) "Statistical Mechanics of Games - Evolutionary Game Theory -," Progress of Theoretical Physics Supplement, No. 179, pp. 216-226

[9] Kikkawa, Mitsuru (2009) "Co-evolution and Diversity in Evolutionary Game Theory : Stochastic Environment," RIMS Kokyuroku, Vol. 1663, pp. 147-152

[10] Losert, Viktor and Ethan Akin (1983) "Dynamics of games and genes: Discrete versus continuous time," Journal of Mathematical Biology, Vol. 17, pp. 241-251

[11] Mehta, Madan Lal, *Random Matrices, Third Edition*, Academic Press, 2004

[12] Metcalfe, J. Stanley, *Evolutionary Economics and Creative Destruction*, Routledge, 1998

[13] Nachbar, John H. (1997) "Prediction, Optimization, and Learning in Repeated Games," Econometrica, Vol. 65, pp. 275-309

[14] Nyarko, Yaw (1998) "Bayesian learning and convergence to Nash equilibria without common priors," Economic Theory, Vol. 11, pp. 643-655

[15] 岡田章，ゲーム理論 新版，有斐閣，2011

[16] Vega-Redondo, Fernando (1997) "The Evolution of Walrasian Behavior," Econometrica, Vol. 65, pp. 375-384